NASA COWEX Data Management Plan

Summary: The Data Management and Field Catalog section from the CASES-99 Operations Plan (http://www.colorado-research.com/cases/C99OpsPlan.htm) was used and modified in the creation of this document. Data from all COWEX participants, including ancillary data, will be made available at different levels:

A quick-look, near real-time data product will be available for determining system performance, data quality, and specifying priority days to focus on when performing data analysis.

A preliminary dataset for use by other COWEX investigators for comparisons and preliminary analyses, but not for publication.

A browse capability using gif or jpeg images to allow other PIs to see what data is available. One or more priority days to focus data analysis efforts in order to speed up publication of results.

A final calibrated dataset. This data will be made publicly available after two years. Many versions of the data may be submitted with the proper documentation.

We propose the following approach for data management. The final objective is a high quality data archive that has easy and timely access by a large community of investigators. This section details the methods we will use as well as necessary guidelines for the PIs to follow in order that this objective may be met. The PI guidelines will include data format conventions, dataset documentation and submission requirements. We will also try to provide supplemental datasets such as, GOES IR, water vapor and infrared images and data, sounding launched from land at such locations as WFF, or other as needed ground sites, etc..

The following data protocols are meant as a guide for all investigators with regard to the processing, quality control, dissemination and sharing of data with other COWEX participants and cooperative projects. The objective of the protocol is to facilitate the timely submission and exchange of high quality data in an open fashion to all interested investigators.

A) All COWEX Science Team (CST) members must agree to submit data within two years after the end of any field campaign to the appropriate archive center (or make available through their own archiving center) to facilitate longer term archival and distribution. This does not mean that no datasets with a better calibration can be submitted after two years. A running log will be kept with the appropriate documentation (see below).

B) Each investigator's data is considered proprietary until the data appear in publication or are published/released via the COWEX archives to the science community. The COWEX archives will be released to the larger science community two years after the campaign ends

C) The CST members may release data to whomever they wish. They may not release data of other COWEX investigators without consent. The direct exchange of data among investigators is encouraged.

D) All COWEX datasets will be considered in the public domain two years after any field mission is completed. This will enhance free and open access to the data. This DOES NOT

remove responsibility from the sponsoring investigator of a given study to solicit the help of the data provider early in the investigation and offer co-authorship when significant intellectual contributions are made to the resulting publication.

E) Datasets submitted to the archive centers must contain sufficient documentation so that users understand the characteristics and attributes of the data and the chosen format. This includes information concerning the quality of the data and may require suitable caveats regarding the data that should be included in any publication using that data.

F) An investigator whose proprietary data are to be used in an investigation has the right to be included among the authors of any resulting publication but must work with the authors to determine such need.

G) CST members publishing COWEX results must always give appropriate acknowledgement and citation of those that collected and provided the data, regardless of contribution to the publication.

H) Any new or derived datasets resulting from collaborative investigations among CST members should be submitted to the COWEX archive center or be made available via links.

I) COWEX will use a distributed archive strategy. Investigators should submit their data to an archive but in special cases may keep their own data collected as part of COWEX so long as they establish appropriate links to this data through the COWEX (to be implemented) web page (or archive center).

J) Certain days will be given priority. As determined and agreed upon by all investigators, certain days may be given priority for data analysis. These days will have measurements of interesting features that fulfill COWEX objectives. Establishing these days as priority is a means of speeding up publication of results.

NASA/Goddard Space Flight Center will be responsible for data management using the following guidelines:

1. Suggest standardized format(s) and guidelines for dataset documentation, status and summary reporting and other important data management procedures as necessary to assure complete documentation of project activities.

2. Provision of a Data Questionnaire to collect information from project PIs regarding their needs for real-time data support in the field as well as information on what datasets they will be providing to the archive both during and after the field season.

3. Collection of supplemental datasets to assist in operations planning and research after any field mission. The actual data collection period (for this ancillary dataset) will start 3-5 days prior to the beginning of the project to assure reliability of the data source and end 3-5 days after

the end of the project. Some of this information will be available in real-time and all will be accessible after the end of any field mission.

4. Archival and distribution of project datasets will be done using the NASA/Goddard Space Flight Center Distributed Active Archive Center (DAAC). All data collected will be made available via the world wide web. For the first two years, data will be available only to COWEX investigators. After two years, the data will be available to the public. COWEX data will be accessible by date, station or location depending on the source and format of the data received. The archivists will continue to work with the project scientists to develop common dataset requirements (e.g. date/time stamps, documentation, forma, etc.) that will simplify access and encourage exchange among the participants and the larger scientific community. The archivists will work with the investigators to assure all data are documented appropriately and the latest version of a given dataset is available as quickly as possible.

Data Management

It is important that the COWEX data management strategy be responsive to the needs of the investigators assuring data are accurate, accessible, well documented and disseminated in a timely fashion.

The first step in organizing COWEX data management support is to understand what data are anticipated from the various components of the program. We will work with the PIs to develop and disseminate a questionnaire for participants that solicit this information from an individual investigator perspective.

It has also been assumed that tasks associated with COWEX data acquisition (e.g. in-field record keeping, backing up field data, data documentation [for catalog purposes], provision of data to data processing locations, processing of raw data into geophysical parameters and initial dataset quality control) will be performed by the participating investigators. The investigators will be requested to document datasets in accordance with the above documentation guidelines so that they could be included in the COWEX catalog in as automated a fashion as possible.

It is recognized that initial field datasets produced by investigators' instrumentation may be in a variety of formats and completeness. It is important that processed data end up in a common format (or documentation describing how to access the data), accessible by all COWEX investigators and eventually the larger scientific community. All datasets should be documented in UTC time (YYYYMMDDHHMMSS) (YYYY=year, MM=month, DD=day, HH=UTC hour, MM=Minute and SS= seconds) .

The COWEX PIs agree to provide data to the archive in either ASCII or netCDF format. Should a PI provide data in a format other than these two, they must provide software or code to read the data.

The importance of providing complete and separate documentation (a readme file) with every COWEX dataset, regardless of format, cannot be over emphasized. It is critical for the long term viability of the comprehensive data base and the easiest way to explain to everyone who might use a dataset important details that might be forgotten in years to come. There are several important components to a complete documentation file that should accompany a given dataset. They include:

Author and/or source of the data
Complete description of the sensors used in the data collection
Complete description of any derived parameters contained in the dataset.
Specific definition of the sample time period (hourly, continuous, etc.) for each variable contained in the dataset.
Specify the units for each measured or derived parameter.
Document the version number.
Minimize the reference to specific dataset file names. They may be modified by the relational database system and file transfer protocol procedures.

A field catalog could be functional during the COWEX field project and would be maintained as necessary to support investigator needs. The catalog would be implemented using a World Wide Web (WWW) interface. Common Web browsers, including Netscape and Internet Explorer, with forms and graphics capabilities would be capable of accessing the catalog. The objective is to have limited data available in near real-time for monitoring instrument performance, system intercomparisons, evaluating completeness of the collected datasets and preliminary analysis. This would become the "living archive" of any field campaign with a quick-look at the data.

The field catalog, may allow on-line data entry (data collection details, field summary notes, certain operational and research data etc.), data browsing (listings, images) and information distribution to other locations worldwide. Daily operations summaries will be prepared that contain as much information about operations (major instrument systems status and sampling times, satellite overpasses, aircraft flight times and tracks, etc.) as desired by the investigators. It is also possible for the project scientists to contribute graphics, i.e. plots in GIF or Postscript format for retention in the catalog. These plots will then be accessible by all participants via Internet. It will be possible to update the status of data collection and instrumentation on a daily basis or more often depending on the platforms.

Supplemental data that will be collected include: GOES IR, Visible and Water vapor images, weather maps (surface maps (every 3 hours), upper air maps (every 12 hours), soundings (every 12 hours), etc). More supplemental data can be collected as requested in the Data Questionnaire.

It is important that all COWEX participants concentrate on post field mission data processing activities to assure timely availability of datasets to all participating investigators as summarized above. Many of these preliminary datasets along with provided operational data should be accessible via the Goddard DAAC as soon as possible after they are received.

These "preliminary" data can be in "native" resolution and format, that is, in the format and resolution the investigators produce from their initial data processing. It is hoped that most preliminary research and operational datasets will become available within 24 months of the end of the field observing program. Following the above protocols, all data will have open access by the COWEX investigators. The final processed dataset (e.g. final format, fully quality controlled, etc.) will replace preliminary data only when directed by the contributing scientists.

Following any field mission the DAAC data management system will be populated with any data provided by the investigators. Data will be searchable by file name and/or data type. The investigators will have complete responsibility for the processing and delivery of their data and documentation to the DAAC within 24 months of the conclusion of the field project. As data are received they will be promptly staged and made available to all COWEX participants. It is not necessary that all COWEX data be housed at the DAAC. Appropriate WWW links can be established so that access is possible at PI home institutions or other data centers. The PIs will provide these links as required.

Information from the field catalog, if implemented, (data collection procedures, instrumentation attributes, graphical products, etc.) during the field season will also be accessible. DAAC would intend to keep the catalog available for several years following the end of the field phase as a ready summary of observations and operations. Links to other data centers or archives will be in place as necessary so that users can have access to all COWEX specific data from a single entry point. It will be possible to make data requests, via World Wide Web (WWW)/Netscape or other interface, and download files via ftp from the catalog or other data center. DAAC will be the archive center for COWEX. The Goddard DAAC system will be used to allow access, perusal and distribution using WWW/Mosaic interface, forms and browse tools. PIs may wish to take on archive responsibilities for certain datasets (i.e. model output, aircraft data, etc.) In that case they will need to coordinate access procedures (specifically, provide a URL link to their archive site) with DAAC.